NONPARAMETRIC ESTIMATION: CUMULANT COMPONENTS IN THE BALANCED ONE-WAY CLASSIFICATION *

D. S. Robson
Cornell University

Abstract

Cumulant components defined for the nonparametric one-way classification describe the manner in which the population departs from a normal mixture of homogeneous normal subpopulations. The unique, minimum variance unbiased estimators of cumulant components are easily determined because of the existence of a complete, sufficient order statistic for the balanced one-way sample. One computational procedure for constructing these estimates, illustrated here with a numerical example, is an algebraic extension of the analysis of variance to the "analysis of cumulants."

*Paper No. 356 from the Department of Plant Breeding, Cornell University.

1. Introduction

The probability model underlying variance component analysis of the one-way classification may be expressed in terms of a linear model

$$X_{ij}=\mu+a_i+e_{ij}$$

and the assumptions concerning the joint distribution of the chance variables a_i and e_{ij}. The usual parametric form of this model has the chance variables a_i and e_{ij} normally and independently distributed with zero means, and with a_i having variance σ_a^2 and the e_{ij} having common variance $\bar{\sigma}_e^2$. The mean μ and the two components σ_a^2 and $\bar{\sigma}_e^2$ of the variance σ_X^2 together represent a sufficient summary description of a mixed population in which these assumptions are known to hold. In the absence of these assumptions the three parameters μ, σ_a^2, and $\bar{\sigma}_e^2$ are no longer definitive; in fact, no finite number of parameters uniquely identifies a nonparametric model for the one-way classification. The significance of variance components as descriptive parameters is therefore considerably reduced when the usual restrictions on the model are removed. Heterogeneity of the "error" variances, for example, is an important and, in practice, common characteristic of mixed populations which is in no sense described by the parameters μ, σ_a^2, and $\bar{\sigma}_e^2$.

The entire sequence of moments or cumulants do, under weak regularity conditions, uniquely characterize a distribution function, and since the mean μ and variance $\sigma_X^2=\sigma_a^2+\bar{\sigma}_e^2$ are already the first two cumulants of X it is natural to examine the components of higher cumulants in looking for additional parameters to describe a mixed population. Furthermore, the convenient properties of the mean and variance derives from the semi-invariance and additive properties of cumulants in general. If the random components of the linear model were, in fact, independent then the υ'th cumulant of X would be the sum of the υ'th cumulants of the random components of X. Components of the higher cumulants may therefore be expected to measure this lack of independence, and so contribute significantly to the description of the probability model. Also, in the usual parametric model with its assumptions of normality and homogeneous variances the components of all cumulants beyond the second vanish; hence, the components of the higher cumulants will, in some sense, measure departure from normality and homogeneity of variance.

2. A nonparametric model for the balanced one-way classification

The "random effects" model for the one-way classification is based upon a probability model described by Feller (1) as a compound experiment. The distribution function $F(x)=\Pr(X \leq x)$ of the chance variable X is regarded as a marginal distribution

$$F(x)=\int_{\mathcal{Y}} F(x \mid y)\,dG(y)$$

where the chance variable Y is called the classification variable. The basic experiment consists of taking an observation y_1 from the population defined by $G(y)$ and then n independent observations $x_{11},\cdots,x_{1n}$ from the population defined by $F(x \mid y_1)$. When c independent repetitions of this basic experiment are made the total outcome forms a one-way array of observations:

y_1	$\cdots$	y_i	$\cdots$	y_c
x_{11}		x_{i1}		x_{c1}
$\vdots$		$\vdots$		$\vdots$
x_{1j}		x_{ij}		x_{cj}
$\vdots$		$\vdots$		$\vdots$
x_{1n}		x_{in}		x_{cn}

In applications, y may ordinarily be regarded as a discrete variable which indexes the subpopulations $F(x|y)=F_y(x)$; consequently, in this non-parametric formulation of the model $G(y)$ is assumed to be a distribution on $\mathcal{Y}=(0,1,2,\cdots,\text{ ad inf})$. The conditional distribution $F(x|y)=F_y(x)$ is assumed to belong either to the class of all absolutely continuous distributions or the class of all discrete distributions for every y. Later, it will be implicitly assumed that the cumulants under discussion do exist.

The linear model for this one-way classification is obtained as an identity. The mean μ_X of the distribution $F(x)$ is

$$\mu_X=\int_{\mathcal{Y}}\mu_X(y)dG(y)$$

where $\mu_X(y)$ is the conditional mean of X,

$$\mu_X(y)=\int_{\mathcal{X}}xdF_y(x)$$

The components of the linear model are defined from the identity

$$X_{ij}\equiv\mu_X+[\mu_X(Y_i)-\mu_X]+[X_{ij}-\mu_X(Y_i)]$$

as

$$a_i=\mu_X(Y_i)-\mu_X$$

$$e_{ij}=X_{ij}-\mu_X(Y_i)$$

The chance variables a_i and e_{ij} defined in this manner obviously have zero means and zero covariance; hence the variance of X_{ij} is expressible as the sum of the two <u>components of variance</u>

$$\sigma_a^2=Ea_i^2=E\left\{E_yX-EX\right\}^2$$

$$\bar{\sigma}_e^2=Ee_{ij}^2=E\left\{\sigma_e^2(Y)\right\} \quad ,$$

where E_y denotes the conditional expectation with respect to $F_y(x)$, and where

$$\sigma_e^2(y)=E_y(X-E_yX)^2$$

is the variance of the distribution $F_y(x)$. Components of the higher cumulants of X also have an interesting and useful interpretation.

3. Population cumulant components

The υ'th moment of $F(x)$, EX^{υ}, will here be denoted by the symbol $\langle\upsilon\rangle$, and the υ'th cumulant by the symbol $[\upsilon]$, after Tukey (6). Cumulants are defined in terms of the moments by the identities

$$\begin{aligned}\varphi_x(t) &= Ee^{tX}\\ &= \sum_{\upsilon=0}^{\infty}\langle\upsilon\rangle\frac{t^{\upsilon}}{\upsilon!}\\ \log\varphi_x(t) &\equiv \sum_{k=1}^{\infty}(-1)^{k-1}\frac{1}{k}(\varphi_x(t)-1)^k\end{aligned}$$

or

$$\sum_{\upsilon=1}^{\infty}[\upsilon]\frac{t^{\upsilon}}{\upsilon!} \equiv \sum_{k=1}^{\infty}(-1)^{k-1}\frac{1}{k}\Big(\sum_{\upsilon=1}^{\infty}\langle\upsilon\rangle\frac{t^{\upsilon}}{\upsilon!}\Big)^k$$

Similarly, the υ'th cumulant of the conditional distribution $F(x|y)$, denoted by $([\upsilon])_y$, is defined in terms of the first υ moments $(\langle 1\rangle)_y,\cdots,(\langle\upsilon\rangle)_y$ of $F(x|y)$ by the identity

$$\log \varphi_{x|y}(t) \equiv \sum_{k=1}^{\infty}(-1)^{k-1}\frac{1}{k}(\varphi_{x|y}(t)-1)^k$$

or

$$\sum_{\upsilon=1}^{\infty}([\upsilon])_y \frac{t^\upsilon}{\upsilon!} \equiv \sum_{k=1}^{\infty}(-1)^{k-1}\frac{1}{k}\Big(\sum_{\upsilon=1}^{\infty}(<\upsilon>)_y \frac{t^\upsilon}{\upsilon!}\Big)^k \tag{3.1}$$

Components of the cumulants $[\upsilon]$ of $F(x)$ are then defined by writing the cumulant generating function of $F(x)$ as

$$\begin{aligned}\log\varphi_x(t) &= \log E\varphi_{x|Y}(t)\\ &= \log Ee^{\log\varphi_{x|Y}(t)}\\ &= \log Ee^{\sum_1^{\infty}([\upsilon])_Y \frac{t^\upsilon}{\upsilon!}}\end{aligned}$$

If $t^\upsilon/\upsilon!$ is, for the moment, written as $s_\upsilon = t^\upsilon/\upsilon!$ then

$$Ee^{\sum_1^{\infty}([\upsilon])_Y s_\upsilon}$$

may be formally identified as the moment generating function of the joint distribution of the chance variables $([1])_Y$, $([2])_Y$, $\cdots$. The cumulant generating function of this joint distribution is then defined by the identity

$$\begin{aligned}&\sum_{\upsilon=1}^{\infty}\sum_{k=1}^{\upsilon}\ \sum_{k_1,\cdots,k_\upsilon \,\Big|\, \substack{\Sigma ik_i=\upsilon \\ \Sigma k_i=k}} [[\underbrace{1,\cdots,1}_{k_1},\cdots \underbrace{\upsilon,\cdots,\upsilon}_{k_\upsilon}]] \frac{s_1^{k_1}\, s_2^{k_2}\cdots s_\upsilon^{k_\upsilon}}{k_1!\, k_2!\cdots k_\upsilon!}\\ (3.2)\quad &\equiv \sum_{h=1}^{\infty}(-1)^{h-1}\frac{1}{h}\Bigg(\sum_{\upsilon=1}^{\infty}\sum_{k=1}^{\upsilon}\ \sum_{k_1,\cdots,k_\upsilon \,\Big|\, \substack{\Sigma ik_i=\upsilon \\ \Sigma k_i=k}}\\ &\qquad E([1])_Y^{k_1}\cdots([\upsilon])_Y^{k_\upsilon}\frac{s_1^{k_1}\, s_2^{k_2}\cdots s_\upsilon^{k_\upsilon}}{k_1!\, k_2!\cdots k_\upsilon!}\Bigg)^h\end{aligned}$$

The doubly bracketed $[[\cdots]]$ terms on the left are called <u>cumulant components</u> since

$$\sum_{\upsilon=1}^{\infty}[\upsilon]\frac{t^{\upsilon}}{\upsilon!}=\sum_{\upsilon=1}^{\infty}\sum_{k=1}^{\upsilon}\sum_{k_1,\cdots,k_\upsilon \Big| \substack{\Sigma i k_i=\upsilon \\ \Sigma k_i=k}}[[\underbrace{1,\cdots,1}_{k_1},\cdots,\underbrace{\upsilon,\cdots,\upsilon}_{k_\upsilon}]]$$

$$\frac{\left(\frac{t^1}{1!}\right)^{k_1}\left(\frac{t^2}{2!}\right)^{k_2}\cdots\left(\frac{t^\upsilon}{\upsilon!}\right)^{k_\upsilon}}{k_1!\;k_2!\;\cdots\;k_\upsilon!}$$

or

$$(3.3)\qquad [\upsilon]=\upsilon!\sum_{k=1}^{\upsilon}\sum_{k_1,\cdots,k_\upsilon \Big| \substack{\Sigma i k_i=\upsilon \\ \Sigma k_i=k}}[[\underbrace{1,\cdots,1}_{k_1},\underbrace{2,\cdots,2}_{k_2},\cdots,\underbrace{\upsilon,\cdots,\upsilon}_{k_\upsilon}]]\frac{1}{\prod_{i}^{\upsilon}k_i!(i!)^{k_i}}$$

For example, the variance of X, denoted here by the symbol

$$[2]=\mathrm{Var}(X)$$

becomes, by (3.3),

$$[2]=2!\left\{[[2]]\frac{1}{0!(1!)^0 1!(2!)^1}+[[1,1]]\frac{1}{2!(1!)^2 0!(2!)^0}\right\}$$

$$=[[2]]+[[1,1]]$$

These two components of $[2]$ are, from (3.2),

$$[[2]]=E([2])_Y$$

$$=E\sigma_e^2(Y)$$

$$=\bar{\sigma}_e^2$$

and

$$[[1,1]] \;=E([1])_Y([1])_Y-E([1])_Y E([1])_Y$$

$$=E(E_Y X)^2-(EX)^2$$

$$=\mathrm{var}\left\{([1])_Y\right\}$$

$$=\sigma_a^2$$

Similarly, the third cumulant of X, conventionally denoted by K_{3X},

$$[3]=E(X-EX)^3=K_{3X},$$

becomes

$$[3]=[[3]]+3[[1,2]]+[[1,1,1]]$$

where, from (3.2),

$$[[3]]=E([3])_Y=EK_{3e}(Y)=\bar{\bar{K}}_{3e}$$

$$\begin{aligned}[[1,2]] &=E([1])_Y([2])_Y-E([1])_Y E([2])_Y\\ &= \mathrm{Cov}\left\{([1])_Y,([2])_Y\right\}\\ &= \mathrm{Cov}\left\{\mu_X(Y),\sigma_e^2(Y)\right\}\end{aligned}$$

$$\begin{aligned}[[1,1,1]] &= E([1])_Y([1])_Y([1])_Y-3E([1])_Y E([1])_Y([1])_Y\\ &\quad +2E([1])_Y E([1])_Y E([1])_Y\\ &= K_3\left\{([1])_Y\right\}\\ &= E(\mu_X(Y)-\mu_X)^3\\ &= K_{3a}\end{aligned}$$

The fourth cumulant of X,

$$[4]=E(X-EX)^4-3(EX^2-(EX)^2)^2=K_{4X},$$

becomes

$$[4]=[[4]]+3[[2,2]]+4[[1,3]]+6[[1,1,2]]+[[1,1,1,1]]$$

where, from (3.2),

$$[[4]]=E([4])_Y=E\,K_{4e}(Y)=\bar{\bar{K}}_{4e}$$

$$[[2,2]]=E([2])^2_Y-(E([2])_Y)^2$$
$$= \text{Var}\left\{([2])_Y\right\}$$
$$= \text{Var}\left\{\sigma^2_e(Y)\right\}$$

$$[[1,3]]=E([1])_Y([3])_Y-E([1])_Y E([3])_Y$$
$$= \text{Cov}\left\{([1])_Y,([3])_Y\right\}$$
$$= \text{Cov}\left\{\mu_X(Y),K_{3e}(Y)\right\}$$

$$[[1,1,2]]=E([1])^2_Y([2])_Y-2E([1])_Y E([1])_Y([2])_Y-E([1])^2_Y E([2])_Y$$
$$+2E([1])_Y E([1])_Y E([2])_Y$$
$$=E(([1])_Y-E([1])_Y)^2([2])_Y-E(([1])_Y-E([1])_Y)^2 E([2])_Y$$
$$= \text{Cov}\left\{(([1])_Y-[1])^2,([2])_Y\right\}$$
$$= \text{Cov}\left\{a^2_Y,\sigma^2_e(Y)\right\}$$

$$[[1,1,1,1]]=E([1])^4_Y-3E([1])^2_Y E([1])^2_Y-4E([1])_Y E([1])^3_Y$$
$$+12E([1])_Y E([1])_Y E([1])^2_Y$$
$$-6E([1])_Y E([1])_Y E([1])_Y E([1])_Y$$
$$=E(([1])_Y-E([1])_Y)^4-3(E([1])^2_Y-E([1])_Y E([1])_Y)^2$$
$$=K_4\left\{([1])_Y\right\}$$
$$=K_{4a}$$

Under the usual assumptions of normality and homogeneity of variance the components of all cumulants of degree greater than 2 vanish. If the subpopulation means are normally distributed then $K_{3a}=K_{4a}=0$; if the subpopulations are homogeneous except for location then $\mathrm{Cov}(\mu_X(Y),\sigma_e^2(Y))=\mathrm{var}(\sigma_e^2(Y))$ $=\mathrm{Cov}(\mu_X(Y),K_{3e}(Y))=\mathrm{Cov}(a_Y^2,\sigma_e^2(Y))=0$; and if the subpopulations are themselves normal then $\bar{K}_{3e}=\bar{K}_{4e}=0$. Otherwise, the components describe the departure from the usual model.

The identification of (3.2) as the cumulant generating function of the joint distribution of the chance variables $([1])_Y$, $([2])_Y$, $([3])_Y$, ..., ad inf. facilitates the interpretation of cumulant compoents. The component

$$[[\underbrace{1,\cdots,1}_{k_1},\underbrace{2,\cdots,2}_{k_2},\cdots,\underbrace{\upsilon,\cdots,\upsilon}_{k_\upsilon}]], \quad \sum_{i=1}^{\upsilon} ik_i=\upsilon, \quad \sum_{i=1}^{\upsilon} k_i=k$$

may then be regarded as a cumulant component of total degree υ in the variate x or, alternatively, as a multivariate cumulant of total degree k in the variates $([1])_y,([2])_y,\cdots,([\upsilon])_y$, being of degree k_1 in the variate $([1])_y$, of degree k_2 in the variate $([2])_y$, and so on. The component $[[r,r,r]]$, for example, may then be interpreted as the third cumulant of the chance variable $([r])_Y$; similarly, $[[r,s]]$, being of degree 1 in the variate $([r])_y$ and of degree 1 in $([s])_y$, and being a bivariate cumulant, may be interpreted as the covariance of the chance variables $([r])_Y$ and $([s])_Y$.

The relationship between cumulant components and the moments is obtained by applying (3.1) to the right side of (3.2), using the abbreviated notation

$$(3.4) \qquad E\left\{(<\upsilon_{11}>)_Y\cdots(<\upsilon_{1r_1}>)_Y\right\}\cdots E\left\{(<\upsilon_{t1}>)_Y\cdots(<\upsilon_{tr_t}>)_Y\right\}$$

$$\underset{\text{def.}}{\equiv} <(<\upsilon_{11},\cdots,\upsilon_{1r_1}>),\cdots,(<\upsilon_{t1},\cdots\upsilon_{tr_t}>)>$$

For example,

$$[[1,2]] = E([1])_Y([2])_Y - E([1])_Y E([2])_Y$$

or, by (3.1),

$$\begin{aligned}
[[1,2]] &= E(\langle 1 \rangle)_Y \left\{ (\langle 2 \rangle)_Y - (\langle 1 \rangle)_Y(\langle 1 \rangle)_Y \right\} \\
&\quad - E(\langle 1 \rangle)_Y E \left\{ (\langle 2 \rangle)_Y - (\langle 1 \rangle)_Y(\langle 1 \rangle)_Y \right\} \\
&= \langle(\langle 1,2 \rangle)\rangle - \langle(\langle 1,1,1 \rangle)\rangle - \langle(\langle 1 \rangle)(\langle 2 \rangle)\rangle \\
&\quad + \langle(\langle 1 \rangle)(\langle 1,1 \rangle)\rangle
\end{aligned}$$

In this manner the following relations are obtained

$$\begin{aligned}
[[1]] &= \langle(\langle 1 \rangle)\rangle \\
[[2]] &= \langle(\langle 2 \rangle)\rangle - \langle(\langle 1,1 \rangle)\rangle \\
[[1,1]] &= \langle(\langle 1,1 \rangle)\rangle - \langle(\langle 1 \rangle)(\langle 1 \rangle)\rangle \\
[[3]] &= \langle(\langle 3 \rangle)\rangle - 3\langle(\langle 1,2 \rangle)\rangle + 2\langle(\langle 1,1,1 \rangle)\rangle \\
[[1,2]] &= \langle(\langle 1,2 \rangle)\rangle - \langle(\langle 1 \rangle)(\langle 2 \rangle)\rangle - \langle(\langle 1,1,1 \rangle)\rangle + \langle(\langle 1 \rangle)(\langle 1,1 \rangle)\rangle \\
[[1,1,1]] &= \langle(\langle 1,1,1 \rangle)\rangle - 3\langle(\langle 1 \rangle)(\langle 1,1 \rangle)\rangle + 2\langle(\langle 1 \rangle)(\langle 1 \rangle)(\langle 1 \rangle)\rangle \\
[[4]] &= \langle(\langle 4 \rangle)\rangle - 3\langle(\langle 2,2 \rangle)\rangle - 4\langle(\langle 1,3 \rangle)\rangle + 12\langle(\langle 1,1,2 \rangle)\rangle \\
&\qquad - 6\langle(\langle 1,1,1,1 \rangle)\rangle \\
[[2,2]] &= \langle(\langle 2,2 \rangle)\rangle - \langle(\langle 2 \rangle)(\langle 2 \rangle)\rangle - 2\langle(\langle 1,1,2 \rangle)\rangle + 2\langle(\langle 1,1 \rangle)(\langle 2 \rangle)\rangle \\
&\qquad + \langle(\langle 1,1,1,1 \rangle)\rangle - \langle(\langle 1,1 \rangle)(\langle 1,1 \rangle)\rangle \\
[[1,3]] &= \langle(\langle 1,2 \rangle)\rangle - \langle(\langle 1 \rangle)(\langle 3 \rangle)\rangle + 3\langle(\langle 1 \rangle)(\langle 1,2 \rangle)\rangle - 3\langle(\langle 1,1,2 \rangle)\rangle \\
&\qquad - 2\langle(\langle 1 \rangle)(\langle 1,1,1 \rangle)\rangle + 2\langle(\langle 1,1,1,1 \rangle)\rangle
\end{aligned}$$

$$[[1,1,2]] = \langle(\langle 1,1,2\rangle)\rangle - 2\langle(\langle 1\rangle)(\langle 1,2\rangle)\rangle - \langle(\langle 1,1\rangle)(\langle 2\rangle)\rangle$$

$$+2\langle(\langle 1\rangle)(\langle 1\rangle)(\langle 2\rangle)\rangle - \langle(\langle 1,1,1,1\rangle)\rangle$$

$$+2\langle(\langle 1\rangle)(\langle 1,1,1\rangle)\rangle + \langle(\langle 1,1\rangle)(\langle 1,1\rangle)\rangle$$

$$-2\langle(\langle 1\rangle)(\langle 1\rangle)(\langle 1,1\rangle)\rangle$$

$$[[1,1,1,1]] = \langle(\langle 1,1,1,1\rangle)\rangle - 3\langle(\langle 1,1\rangle)(\langle 1,1\rangle)\rangle - 4\langle(\langle 1\rangle)(\langle 1,1,1\rangle)\rangle$$

$$+12\langle(\langle 1\rangle)(\langle 1\rangle)(\langle 1,1\rangle)\rangle - 6\langle(\langle 1\rangle)(\langle 1\rangle)(\langle 1\rangle)(\langle 1\rangle)\rangle$$

4. Cumulant component estimation

The limited amount of a priori information contained in a nonparametric model serves to simplify the estimation problem; an estimator which makes "best" use of all available information is relatively easy to find because of the paucity of that information. In the case of a simple sample, $x_1,\cdots,x_m$, for example, the unordered sample $\{x_1,\cdots,x_m\}$, sometimes called the order statistic, is a complete and sufficient statistic if nothing more is known about the population than the fact that it is discrete or continuous. Consequently, any statistic which is a symmetric function of $x_1,\cdots,x_m$ is in this case the unique, minimum variance unbiased estimator of its expected value. This fact, as pointed out by Fraser (2), makes trivial the problem of finding "best" estimators of the moments and polynomial functions of the moments. The minimum variance unbiased estimator of the general term $EX^{\upsilon_1}\, EX^{\upsilon_2} \cdots EX^{\upsilon_r}$, $r \leq m$, of such a polynomial is easily obtained by averaging the unbiased estimator $x_1^{\upsilon_1}\, x_2^{\upsilon_2} \cdots x_r^{\upsilon_r}$ over the m! permutations of $x_1,\cdots,x_m$. The resulting symmetric sample mean, which is the conditional expected value of the estimator $X_1^{\upsilon_1} \cdots X_r^{\upsilon_r}$ given the order statistic $\{x_1,\cdots,x_m\}$ is denoted in Tukey's (6) notation by the primed symbol $\langle \upsilon_1,\cdots,\upsilon_r\rangle'$; thus,

$$\langle \upsilon_1,\cdots,\upsilon_r \rangle' = \frac{1}{(m)_r} \sum_{\substack{0 \le j_1,\cdots,j_r \le m \\ j_1 \ne \cdots \ne j_r}} x_{j_1}^{\upsilon_1}\cdots x_{j_r}^{\upsilon_r}$$

An extension of these simple sample results to the compound type of sample represented by a balanced one-way array $X = \|x_{ij}\|$ makes the estimation of cumulant components correspondingly simple.

4.1 The order statistic of a balanced one-way array

The order statistic $t(X)$, which is a function of the matrix $X = \|x_{ij}\|$ of observations on x, is defined as the set of subsets

$$t(X)=\left\{\left\{x_{11},\cdots,x_{1n}\right\},\cdots,\left\{x_{c1},\cdots,x_{cn}\right\}\right\};$$

that is, the matrices going into $t(X)$ are those obtained by permuting the n elements of each column of X and then permuting the c columns. The order statistic is clearly sufficient for the classes of distributions considered here, and we shall show that it is also complete.

Completeness when c=1 follows directly from the completeness of the order statistic of a simple sample. If the chance variable $h(\left\{X_{11},\cdots,X_{1n}\right\})$ is to have zero expectation for all $F(x)$ considered here then the coefficient of $g(y)=dG(y)$ in

$$Eh(\left\{X_{11},\cdots,X_{1n}\right\})=\sum_{y=0}^{\infty}g(y)E_y h(\left\{X_{11},\cdots,X_{1n}\right\})$$

must vanish for all discrete or absolutely continuous $F(x|y)$. Since, for a fixed y, $\left\{x_{11},\cdots,x_{1n}\right\}$ is complete re all such $F(x|y)$ then $h(\left\{x_{11},\cdots,x_{1n}\right\})$ must be zero everywhere.

Completeness for arbitrary c then follows by induction, for if

$$E_{F(x)}h(t(X)) = \int_{\prod_{i=1}^{c}\prod_{j=1}^{n}\mathcal{X}_{ij}} h(\{\{x_{1j}\},\cdots,\{x_{cj}\}\}) \prod_{i=1}^{c} \sum_{y=0}^{\infty} g(y) \prod_{j=1}^{n} dF(x_{ij}|y)$$

$$= \sum_{y_1=0}^{\infty} g(y_1)^c \int_{\prod_{i=1}^{c-1}\prod_{j=1}^{n}\mathcal{X}_{ij}} \prod_{i=1}^{c-1}\prod_{j=1}^{n} dF(x_{ij}|y_1)$$

$$\left[\int_{\prod_{j=1}^{n}\mathcal{X}_{cj}} h(\{\{x_{1j}\},\cdots,\{x_{cj}\}\}) \prod_{j=1}^{n} dF(x_{cj}|y_1)\right]$$

$$+c \sum_{\substack{y_1,y_2\\ y_1\neq y_2}} g(y_1)^{c-1}g(y_2) \int_{\prod_{i=1}^{c-1}\prod_{j=1}^{n}\mathcal{X}_{ij}} \prod_{i=1}^{c-1}\prod_{j=1}^{n} dF(x_{ij}|y_1)$$

$$\left[\int_{\prod_{j=1}^{n}\mathcal{X}_{cj}} h(\{\{x_{1j}\},\cdots,\{x_{cj}\}\}) \prod_{j=1}^{n} dF(x_{cj}|y_2)\right]$$

$$+\cdots+c! \sum_{\substack{y_1,\cdots,y_c\\ y_1\neq\cdots\neq y_c}} g(y_1)\cdots g(y_c) \int_{\prod_{i=1}^{c-1}\prod_{j=1}^{n}\mathcal{X}_{ij}} \prod_{i=1}^{c-1}\prod_{j=1}^{n} dF(x_{ij}|y_i)$$

$$\left[\int_{\prod_{j=1}^{n}\mathcal{X}_{cj}} h(\{\{x_{1j}\},\cdots,\{x_{cj}\}\}) \prod_{j=1}^{n} dF(x_{cj}|y_c)\right]$$

is zero for all $F(x)$ considered here then the coefficient of $g(y_1)\cdots g(y_c)$, for example, must vanish for all $F(x|y_1),\cdots,F(x|y_c)$. By completeness for $c-1$, however, this implies that for any given $F(x|y_c)$ the integral in the square bracket must vanish; i.e.,

$$\int_{\prod_{j=1}^{n} \mathcal{X}_{cj}} h(\{\{x_{1j}\},\cdots,\{x_{cj}\}\}) \prod_{j=1}^{n} dF(x_{cj}|y_c)=0$$

Since $\{x_{cj}\}=\{x_{c1},\cdots,x_{cn}\}$ is complete re $F(x|y_c)$ then the function $h(\{\{x_{1j}\},\cdots,\{x_{cj}\}\})$ must be zero everywhere.

4.2 Complex symmetric sample means

The sufficiency and completeness of the order statistic $t(X)$ essentially solves the problem of finding the minimum variance unbiased estimators of cumulant components, for if $h(X_{11},\cdots,X_{cn})$ is any unbiased estimator of the cumulant component then $E\{h(X_{11},\cdots,X_{cn})|t(X)\}$ is the unique minimum variance unbiased estimator. Since a cumulant component may be expressed as a linear combination of the moment products defined in (3.4) the estimation problem may be reduced to finding the best estimator of

$$(4.1)\qquad \langle(\langle\upsilon_{11},\cdots,\upsilon_{1r_1}\rangle)\cdots(\langle\upsilon_{t1},\cdots,\upsilon_{tr_t}\rangle)\rangle$$

$$=E\{E_Y X^{\upsilon_{11}}\cdots E_Y X^{\upsilon_{1r_1}}\}\cdots E\{E_Y X^{\upsilon_{t1}}\cdots E_Y X^{\upsilon_{tr_t}}\}$$

If $n<\max(r_1,\cdots,r_t)$ or $c<t$ then no unbiased estimator of (4.1) which is a function of X exists; otherwise, however, $(X_{11}^{\upsilon_{11}}\cdots X_{1r_1}^{\upsilon_{1r_1}})\cdots(X_{t1}^{\upsilon_{t1}}\cdots X_{tr_t}^{\upsilon_{tr_t}})$ is unbiased, and the (unique) minimum variance unbiased estimator is

$$(4.2)\qquad E\{(X_{11}^{\upsilon_{11}}\cdots X_{1r_1}^{\upsilon_{1r_1}})\cdots(X_{t1}^{\upsilon_{t1}}\cdots X_{tr_t}^{\upsilon_{tr_t}})|t(X)\}$$

$$\frac{1}{(n)_{r_1}\cdots(n)_{r_t}(c)_t}\sum_{\substack{0\le i_1,\cdots,i_t\le c\\ i_1\ne\cdots\ne i_t}}\Big(\sum_{\substack{0\le j_1,\cdots,j_{r_1}\le n\\ j_1\ne\cdots\ne j_{r_1}}} x_{i_1j_1}^{\upsilon_{11}}\cdots x_{i_1j_{r_1}}^{\upsilon_{1r_1}}\Big)$$

$$\cdots\Big(\sum_{\substack{0\le j_1,\cdots,jr_t\le n\\ j_1\ne\cdots\ne jr_t}} x_{i_tj_1}^{\upsilon_{t1}}\cdots x_{i_tj_{r_t}}^{\upsilon_{tr_t}}\Big)$$

$$\underset{\text{def}}{\equiv}\langle(\langle\upsilon_{11},\cdots,\upsilon_{1r_1}\rangle)\cdots(\langle\upsilon_{t1},\cdots,\upsilon_{tr_t}\rangle)\rangle'$$

The statistic defined in (4.2) will be called a complex symmetric sample mean of the array X, in contrast to the more easily computed simple symmetric sample mean to be defined later. Clearly, the formula for the (unique) minimum variance unbiased estimator of the cumulant component $[[\upsilon_1,\cdots,\upsilon_r]]$, which will be denoted by the corresponding primed symbol $[[\upsilon_1,\cdots,\upsilon_r]]'$, is then obtainable simply by putting primes on the symbols in the definition of $[[\upsilon_1,\cdots,\upsilon_r]]$ in terms of moment products. For example,

$$[[1]]' = \langle(\langle 1\rangle)\rangle' = \frac{1}{cn}\sum_i\sum_j x_{ij}$$

(4.3) $$[[2]]' = \langle(\langle 2\rangle)\rangle' - \langle(\langle 1,1\rangle)\rangle'$$

$$= \frac{1}{cn}\sum_i\sum_j x_{ij}^2 - \frac{1}{cn(n-1)}\sum_i\sum_{j_1\ne j_2} x_{ij_1}x_{ij_2}$$

(4.4) $$[[1,1]]' = \langle(\langle 1,1\rangle)\rangle' - \langle(\langle 1\rangle)(\langle 1\rangle)\rangle'$$

$$= \frac{1}{cn(n-1)}\sum_i\sum_{j_1\ne j_2} x_{ij_1}x_{ij_2} - \frac{1}{c(c-1)n^2}\sum_{i_1\ne i_2}(\sum_j x_{i_1j})(\sum_j x_{i_2j})$$

4.3 An "analysis of cumulants" method for estimating cumulant components

A more convenient computational form for $[[2]]'$ and $[[1,1]]'$ is available in the familiar analysis of variance method of estimating variance components in a balanced one-way classification. In essence, this procedure consists of computing the sample variance among class means and the average sample variance within classes, equating these two statistics to their expected values in terms of $[[2]]$ and $[[1,1]]$, and solving the resulting equations for the estimates $[[2]]'$ and $[[1,1]]'$. Thus,

$$(4.5) \qquad E\,\frac{\sum_i(\bar{x}_i-\bar{x})^2}{c-1} = \frac{1}{n}([[2]]+n[[1,1]])$$

$$(4.6) \qquad E\,\frac{\sum_i\sum_j(x_{ij}-\bar{x}_i)^2}{c(n-1)} = [[2]]$$

so

$$[[2]]' = \frac{\Sigma\Sigma(x_{ij}-\bar{x}_i)^2}{c(n-1)}$$

$$[[1,1]]' = \frac{\Sigma(\bar{x}_i-\bar{x})^2}{c-1} - \frac{\Sigma\Sigma(x_{ij}-\bar{x}_i)^2}{cn(n-1)}$$

Since the unbiased estimates obtained in this way are also functions of the order statistic $t(X)$ they must, by the completeness of $t(X)$, be identical to the estimates (4.3) and (4.4) given in terms of complex symmetric sample means. It is, in fact, easily verified that

$$[[2]]' = \frac{1}{c(n-1)}\sum_{ij}\Sigma(x_{ij}-\bar{x}_i)^2 = \frac{1}{cn}\sum_{ij}\Sigma x_{ij}^2 - \frac{1}{cn(n-1)}\sum_i \sum_{j_1\neq j_2} x_{ij_1}x_{ij_2}$$

and

$$[[1,1]]' = \frac{1}{c-1}\sum_i(\bar{x}_i-\bar{x})^2 - \frac{1}{cn(n-1)}\sum_{ij}\Sigma(x_{ij}-\bar{x}_i)^2$$

$$= \frac{1}{cn(n-1)}\sum_i \sum_{j_1\neq j_2} x_{ij_1}x_{ij_2} - \frac{1}{c(c-1)n^2}\sum_{i_1\neq i_2}(\sum_j x_{i_1j})(\sum_j x_{i_2j})$$

The statistics on the left hand sides of (4.5) and (4.6) may be conveniently denoted by the symbols $[[1]'[1]']'$ and $[[2]']'$, and their expectations by $[[1]'[1]']$ and $[[2]']$, respectively, in the sense that $[[1]'[1]']$ is the second cumulant (variance) of the chance variable

$$([1]')_{Y_i} = \frac{1}{n}\sum_{j=1}^{n} x_{ij} = \bar{x}_i$$

and $[[2]']$ is the first cumulant (mean) of the chance variable

$$([2]')_{Y_i} = \frac{1}{n-1}\sum_{j=1}^{n}(x_{ij}-\bar{x}_i)^2 = s_i^2$$

Similarly,

$$[[1]'[1]']' = \frac{1}{c-1}\Sigma(\bar{x}_i-\bar{x})^2$$

is the second sample cumulant of $([1]')_{Y_i} = \bar{x}_i$, and

$$[[2]']' = \frac{1}{c}\sum_i \frac{\sum_j(x_{ij}-\bar{x}_i)^2}{n-1}$$

is the first sample cumulant of $([2]')_{Y_i} = s_i^2$.

The computational simplicity of the analysis of variance method of estimating components of the second cumulant suggests that an extension of the method might provide a simple way of estimating components of the higher cumulants. The obvious extension of the approach taken above consists of expressing the multivariate cumulants

$$[\underbrace{[1]',\cdots,[1]'}_{k_1},\underbrace{[2]',\cdots,[2]'}_{k_2},\cdots,\underbrace{[\upsilon]',\cdots,[\upsilon]'}_{k_\upsilon}] \;;\; \Sigma ik_i=\upsilon \quad (\upsilon\le n)$$
$$1\le \Sigma k_i \le \upsilon$$

of the chance variables $([1]')_Y$, $([2]')_Y$, $\cdots$, $([\upsilon]')_Y$ in terms of the complex cumulant components to be estimated, first estimating these multivariate cumulants and then, by a subtraction process estimating the complex cumulant components. For $\upsilon=3$, for example, this approach will be seen to give the results

$$[[1]'[1]'[1]'] = \frac{1}{n^2}\left\{[[3]]+3n[[1,2]]+n^2[[1,1,1]]\right\}$$

$$[[1]'[1]'[1]']' = \frac{c}{(c-1)(c-2)} \sum_i(\bar{x}_i-\bar{x})^3$$

$$[[1]'[2]'] = \frac{1}{n}\left\{[[3]]+n[[1,2]]\right\}$$

$$[[1]'[2]']' = \frac{1}{(c-1)(n-1)} \sum_i(\bar{x}_i-\bar{x})\sum_j(x_{ij}-\bar{x}_i)^2$$

$$[[3]'] = [[3]]$$

$$[[3]']' = \frac{n}{c(n-1)(n-2)} \sum_i\sum_j(x_{ij}-\bar{x}_i)^3$$

from which the estimates $[[3]]'$, $[[1,2]]'$, and $[[1,1,1]]'$ may be obtained by an obvious subtraction process. Again, since the sample multivariate cumulants on the left hand sides of these equations are themselves functions of the order statistic $t(X)$ then the estimates of the cumulant components computed in this way are the unique, minimum variance unbiased estimators.

The relationship between the multivariate cumulants of $([1]')_Y,\cdots,([n]')_Y$ and the cumulant components may be determined from the multivariate cumulant generating function

$$(4.7)\qquad \sum_{\upsilon=1}^{\infty}\sum_{k=1}^{\upsilon}\sum_{k_1,\cdots,k_n\,\middle|\,\substack{\Sigma ik_i=\upsilon\\ \Sigma k_i=k}}[\underbrace{[1]',\cdots,[1]'}_{k_1},\cdots,\underbrace{[n]',\cdots,[n]'}_{k_n}]\ \frac{t_1^{k_1}\cdots t_n^{k_n}}{k_1!\cdots k_n!}$$

$$\equiv \log Ee^{\sum_{\upsilon=1}^{n}([\upsilon]')_Y t_\upsilon}$$

$$\equiv \sum_{h=1}^{\infty}(-1)^{h-1}\frac{1}{h}\left\{\sum_{\upsilon=1}^{\infty}\sum_{k=1}^{\upsilon}\sum_{k_1,\cdots,k_n\,\middle|\,\substack{\Sigma ik_i=\upsilon\\ \Sigma k_i=k}} E([1]')_Y^{k_1}\cdots([n]')_Y^{k_n}\ \frac{t_1^{k_1}\cdots t_n^{k_n}}{k_1!\cdots k_n!}\right\}^h$$

and the identity (3.2). For $1\leq\upsilon\leq n$, the estimable range of υ, a 1-1 correspondence exists between the terms of (3.2) and those of (4.7); namely, on the left hand sides

$$[[\underbrace{1,\cdots,1}_{k_1},\cdots,\underbrace{\upsilon,\cdots,\upsilon}_{k_\upsilon}]]\rightarrow[\underbrace{[1]',\cdots,[1]'}_{k_1},\cdots,\underbrace{[\upsilon]',\cdots,[\upsilon]'}_{k_\upsilon}]$$

and on the right hand sides

$$E([1])_Y^{k_1}\cdots([\upsilon])_Y^{k_\upsilon}\ \longrightarrow\ E([1]')_Y^{k_1}\cdots([1]')_Y^{k_\upsilon}$$

Consequently, the relations derived from (3.2) induce corresponding relations derivable from (4.7); for example, the definition

$$[[1,2]]=E([1])_Y([2])_Y-E([1])_YE([2])_Y$$

derived from (3.2) induces the relation

$$[[1]'[2]']=E([1]')_Y([2]')_Y-E([1]')_YE([2]')_Y$$

The first term on the right,

$$E([1]')_Y([2]')_Y = E\left\{E\left\{([1]')_Y([2]')_Y \middle| Y\right\}\right\},$$

may be written

$$E\left\{E\left\{([1]')_Y([2]')_Y \middle| Y\right\}\right\} = E\left\{\frac{1}{n}([3])_Y + ([1])_Y([2])_Y\right\}$$

and the second term on the right is clearly

$$E([1]')_Y E([2]')_Y = E([1])_Y E([2])_Y$$

Consequently, the covariance of $([1]')_Y$ and $([2]')_Y$ becomes

$$[[1]',[2]'] = \frac{1}{n}E([3])_Y + E([1])_Y([2])_Y - E([1])_Y E([2])_Y$$

$$= \frac{1}{n}[[3]] + [[1,2]]$$

as indicated earlier. A procedure for evaluating $E\left\{([1]')_Y^{k_1}\cdots([n]')_Y^{k_n} \middle| Y\right\}$ in terms of the cumulants of $F(x|Y)$ is given in a paper by Robson (5) on the multivariate cumulants of a simple sample.

Application of this method to the fourth cumulant results in the following analysis

$$[[1]',[1]',[1]',[1]'] = \frac{1}{n^3}\left\{[[4]] + 3n[[2,2]] + 4n[[1,3]] + 6n^2[[1,1,2]] + n^3[[1,1,1,1]]\right\}$$

$$[[1]'[1]'[1]'[1]']' = \frac{c}{(c-1)(c-2)(c-3)}\left\{(c+1)\sum_i(\bar{x}_i-\bar{x})^4 - \frac{3(c-1)}{c}\left(\sum_i(\bar{x}_i-\bar{x})^2\right)^2\right\}$$

$$[[1]'[1]'[2]'] = \frac{1}{n^2}\Big\{[[4]]+n[[2,2]]+2n[[1,3]]+n^2[[1,1,2]]\Big\}$$

$$[[1]'[1]'[2]']' = \frac{c}{(c-1)(c-2)(n-1)}\sum_i\Big\{(\bar{x}_i-\bar{x})^2 - \frac{1}{c}\sum_i(\bar{x}_i-\bar{x})^2\Big\}\sum_j(x_{ij}-\bar{x}_i)^2$$

$$[[2]'[2]'] = \frac{1}{n}\Big\{[[4]]+n[[2,2]]+\frac{2n}{n-1}E([2])_Y([2])_Y\Big\}$$

$$[[2][2]]' = \frac{1}{(c-1)(n-1)^2}\Big\{\sum_i(\sum_j(x_{ij}-\bar{x}_i)^2)^2 - \frac{1}{c}\sum_i(\sum_j(x_{ij}-\bar{x}_i)^2)^2\Big\}$$

$$[[1]'[3]'] = \frac{1}{n}\Big\{[[4]]+n[[1,3]]\Big\}$$

$$[[1]'[3]']' = \frac{n}{(c-1)(n-1)(n-2)}\sum_i(\bar{x}_i-\bar{x})\sum_j(x_{ij}-\bar{x}_i)^3$$

$$[[4]']' = [[4]]'$$

$$= \frac{n}{c(n-1)(n-2)(n-3)}\Big\{(n+1)\sum_{ij}\sum(x_{ij}-\bar{x}_i)^4 - \frac{3(n-1)}{n}\sum_i(\sum_j(x_{ij}-\bar{x}_i)^2)^2\Big\}$$

Notice that $[[2]'[2]']$ contains the term $E([2])_Y^2$, which is not a component of the fourth cumulant; this term may be estimated by including the additional equation

$$\frac{1}{n+1}E\Big\{\frac{c-1}{c}[[2]'[2]']'+[[2]']'[[2]']'-[[4]]'\Big\} = \frac{1}{n-1}E([2])_Y^2$$

The presence of this term detracts from the usefulness and convenience of the method, and the corresponding setup for the fifth and higher cumulants is further encumbered by the presence of terms which are not cumulant components. Its appearance here could be anticipated since $[[2]'[2]']$ is the variance of the statistic $([2]')_Y$ and is therefore non-zero, even in the usual normal, parametric model where all components of the fourth cumulant vanish. The "extra" term in $[[2]'[2]']$ is, of course, the variance of $([2])_Y$ in the homogeneous normal case.

4.4 Simple symmetric sample means: computing formulas

Neither the complex symmetric sample means of section 4.2 nor the sample multivariate cumulants of preceding section are in a form which can be conveniently fed into a computer. The estimate $[[2]]'$, for example, is given in the two forms

$$[[2]]' = \frac{1}{cn}\sum_{ij}\sum x_{ij}^2 - \frac{1}{cn(n-1)}\sum_i \sum_{j_1 \neq j_2} x_{ij_1}x_{ij_2}$$

and

$$[[2]]' = \frac{1}{c(n-1)}\sum_{ij}\sum (x_{ij}-\bar{x}_i)^2 ,$$

respectively, while the desired form for computing purposes is

$$[[2]]' = \frac{1}{c(n-1)}\left\{\sum_{ij}\sum x_{ij}^2 - \frac{1}{n}\sum_i(\sum_j x_{ij})^2\right\}$$

or

$$[[2]]' = \frac{n}{n-1}\left\{\frac{1}{c}\sum_i \frac{1}{n}\sum_j x_{ij}^2 - \frac{1}{c}\sum_i(\frac{1}{n}\sum_j x_{ij})^2\right\}$$

The terms in the latter expression are called simple symmetric means of X and are denoted by angle bracket symbols as follows

$$\langle\langle \nu_1\rangle' \cdots \langle \nu_r\rangle'\rangle' \underset{\text{def}}{\equiv} \frac{1}{cn^r}\sum_{i}^{c}(\sum_{j}^{n} x_{ij}^{\nu_1})\cdots(\sum_{j}^{n} x_{ij}^{\nu_r}) .$$

In this notation the computing formula for $[[2]]'$ becomes

$$[[2]]' = \frac{n}{n-1}\left\{\langle\langle 2\rangle'\rangle' - \langle\langle 1\rangle'\langle 1\rangle'\rangle'\right\}$$

A simple relationship exists between the simple and complex symmetric means of X, so that any linear function of complex symmetric means, such as

$$[[2]]' = \langle(\langle 2\rangle)\rangle' - \langle(\langle 1,1\rangle)\rangle'$$

is easily converted into a computing formula in terms of simple symmetric means. This relation is given by Robson (5) for the simple sample case (c=1 or n=1) and may be applied to present case in two stages. The notation required to produce a general formula is somewhat cumbersome even though the relation is relatively simple. To obtain the expression for the complex symmetric mean $\langle(\langle \upsilon_{11},\cdots,\upsilon_{1r_1}\rangle),\cdots,(\langle \upsilon_{t1},\cdots,\upsilon_{tr_t}\rangle)\rangle'$ in terms of simple symmetric means, Robson's formula is applied first to the elements $(\cdots),\cdots,(\cdots)$ and second to the $\langle\cdots\rangle$ within each $(\cdots)$. Let I_t denote the set $\{1,2,\cdots,t\}$, let $I_t(k_i)$ denote a collection of k_i disjoint subsets of I_t containing i elements each, and let $I_t(k_1,\cdots,k_t;k)$ denote a collection $\{I_t(k_1),\cdots,I_t(k_t)\}$ in which $\Sigma k_i=k$ and in which every element of I_t appears exactly once, so that $\Sigma ik_i=t$. Then as the first stage,

$$\langle(\langle \upsilon_{11},\cdots,\upsilon_{1r_1}\rangle),\cdots,(\langle \upsilon_{t1},\cdots,\upsilon_{tr_t}\rangle)\rangle'$$

$$= \frac{1}{(c)_t}\sum_{k=1}^{t}(-1)^{t-k}c^k \sum_{k_1,\cdots,k_t \,\middle|\, \substack{\Sigma ik_i=t \\ \Sigma k_i=k}} \prod_{i=1}^{t}(i-1)!^{k_i}$$

$$\sum_{\{I_t(k_1,\cdots,k_t;k)\}} \prod_{i=1}^{t} \prod_{\{j_1,\cdots,j_i\}\in I_t(k_i)} \langle\langle \upsilon_{j_1 1},\cdots,\upsilon_{j_1 r_{j_1}}\rangle' \cdots \langle \upsilon_{j_i 1},\cdots,\upsilon_{j_i r_{j_i}}\rangle'\rangle'$$

The second stage consists of applying Robson's simple sample formula directly to each factor $\langle\upsilon_{i1},\cdots,\upsilon_{ir_i}\rangle'$ in every term. Specifically,

$$\langle\upsilon_{i1},\cdots,\upsilon_{ir_i}\rangle' = \frac{1}{(n)_{r_i}} \sum_{k=1}^{r_i} (-1)^{r_i-k} n^k \sum_{k_1,\cdots,k_{r_i} \left| \substack{\Sigma\upsilon k_\upsilon = r_i \\ \Sigma k_\upsilon = k} \right.} \prod_{\upsilon=1}^{r_i} (\upsilon-1)!^{k_\upsilon}$$

$$\sum_{\{I_{r_i}(k_1,\cdots,k_{r_i};k)\}} \prod_{\upsilon=1}^{r_i} \prod_{\{j_1,\cdots,j_\upsilon\}\in I_{r_i}(k_\upsilon)} \langle \upsilon_{ij_1}+\cdots+\upsilon_{ij_\upsilon}\rangle'$$

For example, the two stages in constructing the computing formula for $\langle(\langle\upsilon_{11}\rangle),(\langle\upsilon_{21},\upsilon_{22}\rangle),(\langle\upsilon_{31},\upsilon_{32},\upsilon_{33}\rangle)\rangle'$ are, first

$$\frac{1}{(c)_3}\Big\{c(2!)\langle\langle\upsilon_{11}\rangle'\langle\upsilon_{21},\upsilon_{22}\rangle'\langle\upsilon_{31},\upsilon_{32},\upsilon_{33}\rangle'\rangle'$$

$$-c^2(\langle\langle\upsilon_{11}\rangle'\rangle'\langle\langle\upsilon_{21},\upsilon_{22}\rangle'\langle\upsilon_{31},\upsilon_{32},\upsilon_{33}\rangle'\rangle'$$

$$+\langle\langle\upsilon_{21},\upsilon_{22}\rangle'\rangle'\langle\langle\upsilon_{11}\rangle'\langle\upsilon_{31},\upsilon_{32},\upsilon_{33}\rangle'\rangle'$$

$$+\langle\langle\upsilon_{31},\upsilon_{32},\upsilon_{33}\rangle'\rangle'\langle\langle\upsilon_{11}\rangle'\langle\upsilon_{21},\upsilon_{22}\rangle'\rangle')$$

$$+c^3\langle\langle\upsilon_{11}\rangle'\rangle'\langle\langle\upsilon_{21},\upsilon_{22}\rangle'\rangle'\langle\langle\upsilon_{31},\upsilon_{32},\upsilon_{33}\rangle\rangle'\Big\}$$

and, second, replacing the factors $\langle\upsilon_{11}\rangle'$, $\langle\upsilon_{21},\upsilon_{22}\rangle'$, and $\langle\upsilon_{31},\upsilon_{32},\upsilon_{33}\rangle'$ by

$$\langle \upsilon_{11} \rangle' = \langle \upsilon_{11} \rangle'$$

$$\langle \upsilon_{21}, \upsilon_{22} \rangle' = \frac{1}{n^2} \left\{ -n \langle \upsilon_{21}+\upsilon_{22} \rangle' + n^2 \langle \upsilon_{21} \rangle' \langle \upsilon_{22} \rangle' \right\}$$

$$\langle \upsilon_{31}, \upsilon_{32}, \upsilon_{33} \rangle' = \frac{1}{n^3} \left\{ n(2!) \langle \upsilon_{31}+\upsilon_{32}+\upsilon_{33} \rangle' - n^2 (\langle \upsilon_{31} \rangle' \langle \upsilon_{32}+\upsilon_{33} \rangle' + \langle \upsilon_{32} \rangle' \langle \upsilon_{31}+\upsilon_{33} \rangle' + \langle \upsilon_{33} \rangle' \langle \upsilon_{31}+\upsilon_{32} \rangle') + n^3 \langle \upsilon_{31} \rangle' \langle \upsilon_{32} \rangle' \langle \upsilon_{33} \rangle' \right\}$$

Then, for example,

$$\langle \langle \upsilon_{11} \rangle' \langle \upsilon_{21}, \upsilon_{22} \rangle' \rangle' = \frac{1}{n^2} \left\{ -n \langle \langle \upsilon_{11} \rangle' \langle \upsilon_{21}+\upsilon_{22} \rangle' \rangle' + n^2 \langle \langle \upsilon_{11} \rangle' \langle \upsilon_{21} \rangle' \langle \upsilon_{22} \rangle' \rangle' \right\}$$

Applied to the analysis of variance, this formula gives

$$[[1]'[1]']' = \frac{1}{n} \left\{ [[2]]' + n[[1,1]]' \right\}$$

$$= \frac{1}{n} \left\{ (\langle(\langle 2 \rangle)\rangle' - \langle(\langle 1,1 \rangle)\rangle') + n(\langle(\langle 1,1 \rangle)\rangle' - \langle(\langle 1 \rangle)(\langle 1 \rangle)\rangle') \right\}$$

$$= \frac{1}{n} \left\{ \langle\langle 2 \rangle'\rangle' + \frac{n-1}{n(n-1)}(-n \langle\langle 2 \rangle'\rangle' + n^2 \langle\langle 1 \rangle'\langle 1 \rangle'\rangle') - \frac{n}{c(c-1)}(-c \langle\langle 1 \rangle'\langle 1 \rangle'\rangle' + c^2 \langle\langle 1 \rangle'\rangle' \langle\langle 1 \rangle'\rangle') \right\}$$

$$= \frac{c}{c-1}(\langle\langle 1 \rangle'\langle 1 \rangle'\rangle' - \langle\langle 1 \rangle'\rangle' \langle\langle 1 \rangle'\rangle')$$

$$[[2]]' = \langle(\langle 2 \rangle)\rangle' - \langle(\langle 1,1 \rangle)\rangle'$$

$$= \langle\langle 2 \rangle'\rangle' - \frac{1}{n(n-1)}(-n \langle\langle 2 \rangle'\rangle' + n^2 \langle\langle 1 \rangle' \langle 1 \rangle'\rangle')$$

$$= \frac{n}{n-1}(\langle\langle 2 \rangle'\rangle' - \langle\langle 1 \rangle' \langle 1 \rangle'\rangle')$$

Similarly, the computing formulas for the third and fourth cumulant component analyses are

$$
\begin{aligned}
[[1]'[1]'[1]']' &= \frac{c^2}{(c-1)(c-2)}\Big(\langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle' - 3\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 1\rangle'\rangle' \\
&\quad + 2\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\rangle'\Big)
\end{aligned}
$$

$$
\begin{aligned}
[[1]'[2]']' &= \frac{cn}{(c-1)(n-1)}\Big(\langle\langle 1\rangle'\langle 2\rangle'\rangle' - \langle\langle 1\rangle'\rangle'\langle\langle 2\rangle'\rangle' \\
&\quad - \langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle' + \langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 1\rangle'\rangle'\Big)
\end{aligned}
$$

$$
\begin{aligned}
[[3]']' &= \frac{n^2}{(n-1)(n-2)}\Big(\langle\langle 3\rangle'\rangle' - 3\langle\langle 1\rangle'\langle 2\rangle'\rangle' \\
&\quad + 2\langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle'\Big)
\end{aligned}
$$

and

$$
\begin{aligned}
[[1]'[1]'[1]'[1]']' &= \frac{c^3}{(c-1)(c-2)(c-3)}\Big(\langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle' \\
&\quad - 3\langle\langle 1\rangle'\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 1\rangle'\rangle' \\
&\quad - 4\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle' \\
&\quad + 12\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 1\rangle'\rangle' \\
&\quad - 6\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\rangle'\Big) \\
&\quad + \frac{c^2}{(c-1)(c-2)(c-3)}\Big(\langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle' \\
&\quad + 3\langle\langle 1\rangle'\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 1\rangle'\rangle' \\
&\quad - 4\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle'\Big)
\end{aligned}
$$

$$[[1]'[1]'[2]']' = \frac{c^2n^2}{(c-1)(c-2)(n-1)}(\langle\langle 1\rangle'\langle 1\rangle'\langle 2\rangle'\rangle'$$

$$-2\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 2\rangle'\rangle' - \langle\langle 1\rangle'\langle 1\rangle'\rangle'\langle\langle 2\rangle'\rangle'$$

$$+2\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\rangle'\langle\langle 2\rangle'\rangle' - \langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle'$$

$$+2\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle'$$

$$+\langle\langle 1\rangle'\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 1\rangle'\rangle'$$

$$-2\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 1\rangle'\rangle')$$

$$[[2]'[2]']' \quad - \frac{2}{n+1}\left\{\frac{c-1}{c}[[2]'[2]']' + [[2]']'[[2]']' - \frac{1}{n}[[4]']'\right\}$$

$$= \frac{n^2}{(n-1)}\left(\frac{1}{(c-1)(n-1)} + \frac{n-6}{(n-2)(n-3)}\right)(\langle\langle 2\rangle'\langle 2\rangle'\rangle'$$

$$-2\langle\langle 1\rangle'\langle 1\rangle'\langle 2\rangle'\rangle' + \langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle')$$

$$-\frac{cn^2}{(c-1)(n-1)^2}(\langle\langle 2\rangle'\rangle'\langle\langle 2\rangle'\rangle' - 2\langle\langle 1\rangle'\langle 1\rangle'\rangle'\langle\langle 2\rangle'\rangle'$$

$$+\langle\langle 1\rangle'\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 1\rangle'\rangle')$$

$$+\frac{2n}{(n-1)(n-2)(n-3)}(\langle\langle 4\rangle'\rangle' + 3\langle\langle 2\rangle'\langle 2\rangle'\rangle' - 4\langle\langle 1\rangle'\langle 3\rangle'\rangle')$$

which has expectation $\frac{1}{n}\left\{[[4]] + n[[2,2]]\right\}$,

$$[[1][3]]' \quad = \frac{cn^2}{(c-1)(n-1)(n-2)}(\langle\langle 1\rangle'\langle 3\rangle'\rangle' - \langle\langle 1\rangle'\rangle'\langle\langle 3\rangle'\rangle'$$

$$+3\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 2\rangle'\rangle'$$

$$-3\langle\langle 1\rangle'\langle 1\rangle'\langle 2\rangle'\rangle' - 2\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle'$$

$$+2\langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle')$$

$$[[4]]' = \frac{n^3}{(n-1)(n-2)(n-3)}(\langle\langle 4\rangle'\rangle' - 3\langle\langle 2\rangle'\langle 2\rangle'\rangle' - 4\langle\langle 1\rangle'\langle 3\rangle'\rangle'$$

$$+12\langle\langle 1\rangle'\langle 1\rangle'\langle 2\rangle'\rangle' - 6\langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle')$$

$$+\frac{n^2}{(n-1)(n-2)(n-3)}(\langle\langle 4\rangle'\rangle' + 3\langle\langle 2\rangle'\langle 2\rangle'\rangle' - 4\langle\langle 1\rangle'\langle 3\rangle'\rangle')$$

A numerical illustration using these computing formulas is given in the Appendix.

5. Discussion

Until recently, cumulants beyond the second have found little application outside of their occasional use in fitting Pearson Type distribution functions. Tukey's recent introduction of the polykay system, which simplifies both the numerical and algebraic computation problem, has served to place new emphasis on the cumulants, opening up many interesting possibilities for advancing the theory of survey sampling and, in general, enlarging our view of the nonparametric estimation problem.

Variance component estimation, a technique of increasing practical importance in such fields as plant and animal breeding, has been the topic most intensively studied in terms of the polykay system. The polykay approach to variance component estimation is nonparametric in the sense that no assumptions are made concerning the functional form of the underlying distribution functions, other than the assumption of the existence of moments. In the case of finite populations, of course, the required moments always exist. Since the problem is viewed nonparametrically then variance component estimates actually represent only a small fraction of the information contained in the sample, and while it is certainly desirable to estimate moments of the sampling distribution of the variance component estimates (Hooke (3), Robson (4), Tukey (7),(8)) it would seem even more desirable to direct this computing effort toward a further description of the population itself. The polykay approach and the modern computing machinery now make practicable the estimation of higher cumulant components and, therefore, the extraction of additional information from the sample.

The present results for the balanced one-way classification suggest numerous possible extensions and related problems. Work is now in progress, for example, on constructing an algebra to simplify the manipulation of cumulant components and symmetric means of compound type distributions; the balanced nested and r-way classifications are under study for both the univariate and multivariate case, and the possibility of developing a genetic cumulant component analysis as an extension of the genetic variance component concept is being investigated.

The form of the analysis presented in section 4.3, representing an extension of the analysis of variance form, also suggests the possibility of constructing test procedures for testing the assumptions underlying the analysis of variance. The invariance properties of the mean square ratio $n[[1]'[1]']'/[[2]']'$ apply also to a ratio such as

$$\frac{n[[1]'[2]']'}{[[3]]'} = \frac{\sum_i(\bar{\bar{x}}_i-\bar{\bar{x}})\sum_j(x_{ij}-\bar{x}_i)^2}{\sum\sum_{ij}(x_{ij}-\bar{\bar{x}}_i)^3} = \frac{[[3]]'+n[[1,2]]'}{[[3]]'}$$

implying that under the usual assumptions of normality and homogeneous variances the distribution of this ratio depends only upon the known parameters c and n. If this distribution could be tabulated then similar ratios from a balanced two-way sample would also provide various tests of additivity.

The unbalanced classifications present a much more difficult estimation problem because of the apparent absence of a complete sufficient statistic. Completeness in the balanced case is an extremely useful property, eliminating the necessity of comparing different unbiased estimators of the same parameter, which becomes a problem of major concern in the unbalanced case. It is interesting to note that the completeness demonstrated here implies that the unbiased estimates of the moments of variance component estimates in balanced classifications given previously (Tukey (7) and Robson (4)) are, in fact, the minimum variance unbiased estimators.

References

(1) William Feller, An Introduction to Probability Theory and its Applications, first edition (1950), John Wiley and Sons, Inc., New York.

(2) D. A. S. Fraser, Nonparametric Methods in Statistics (1957), John Wiley and Sons, Inc., New York.

(3) Robert Hooke, "Symmetric functions of a two-way array", Annals of Mathematical Statistics, 27 (1956), 55-79.

(4) Douglas S. Robson, "Applications of the k_4 statistic to genetic variance component analyses", Biometrics, (1956), 433-444.

(5) ________________, "Some applications of multivariate polykays to the theory of unbiased ratio-type estimation", Journal of the American Statistical Association, 52, (1957), 511-522.

(6) John W. Tukey, "Keeping moment-like sampling computations simple", Annals of Mathematical Statistics, 27, (1956), 37-54.

(7) ____________, "Variances of variance components: I. Balanced designs", Annals of Mathematical Statistics, 27, (1956), 722-736.

(8) ____________, "Variances of variance components: III. Third moments in a balanced single classification", Annals of Mathematical Statistics, 28, (1957), 378-384.

Appendix

The computational procedure for estimating the components of the first four cumulants in a balanced one-way classification is illustrated here with a numerical example. The sample in Table 1 was drawn from a population satisfying the model II analysis of variance assumptions concerning the distributional properties of the components of the linear model $X_{ij}=\mu+a_i+e_{ij}$. The a_i, $i=1,\cdots,20$, were obtained as a random sample of 20 observations from the standard normal distribution and, for each i, the e_{ij}, $j=1,\cdots,10$, were obtained as independent random samples of size 10 from the standard normal distribution; μ was taken as $\mu=5$ in order to eliminate negative observations. The components of the first two cumulants are then $[1]=5$, $[[1,1]]=1$, $[[2]]=1$, and all other cumulant components are zero.

Simple symmetric sample means

$$\langle\langle 1\rangle'\rangle' = \frac{1}{cn}\,\Sigma\Sigma x = \frac{1008.7}{200} = 5.0435$$

$$\langle\langle 2\rangle'\rangle' = \frac{1}{cn}\,\Sigma\Sigma x^2 = \frac{5581.25}{200} = 27.90625$$

$$\langle\langle 1\rangle'\langle 1\rangle'\rangle' = \frac{1}{cn^2}\,\Sigma(\Sigma x)^2 = \frac{53925.39}{2000} = 26.962695$$

$$\langle\langle 3\rangle'\rangle' = \frac{1}{cn}\,\Sigma\Sigma x^3 = \frac{33094.825}{200} = 165.474125$$

$$\langle\langle 1\rangle'\langle 2\rangle'\rangle' = \frac{1}{cn^2}\,\Sigma(\Sigma x)(\Sigma x^2) = \frac{311581.077}{2000} = 155.7905385$$

$$\langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle' = \frac{1}{cn^3}\,\Sigma(\Sigma x)^3 = \frac{3021584.191}{20000} = 151.07920955$$

Table 1. A one-way sample array from a normal mixture of normal subpopulations

(ii)

Sample Number i\j	Sample Observations x_{ij} 1	2	3	4	5	6	7	8	9	10	Σx	$(\Sigma x)^2$	Σx^2	Σx^3	Σx^4
1	7.2	5.6	6.2	4.1	7.6	5.0	5.6	7.0	4.4	6.9	59.6	3,552.16	368.54	2,352.398	15,418.2338
2	5.8	4.9	7.1	6.4	6.9	7.1	7.4	6.0	6.9	7.6	66.1	4,369.21	443.17	3,007.945	20,632.4869
3	6.1	6.6	6.0	5.7	5.9	6.3	5.6	5.8	8.2	4.0	60.2	3,624.04	372.00	2,357.192	15,313.0068
4	6.3	7.5	5.3	6.6	8.4	6.6	8.4	6.5	5.7	7.0	68.3	4,664.89	476.01	3,384.017	24,522.4437
5	3.0	3.6	4.4	4.0	6.6	4.0	4.4	3.4	3.1	3.9	40.4	1,632.16	172.62	787.934	3,865.3842
6	2.6	3.2	2.2	2.5	3.3	3.3	3.3	0.6	3.2	3.2	27.4	750.76	81.60	250.180	778.6644
7	3.3	4.6	2.3	5.6	5.0	5.0	3.8	3.4	3.9	5.2	42.1	1,772.41	186.95	865.159	4,132.4243
8	4.9	5.0	2.3	4.3	4.6	4.3	5.3	4.0	3.9	5.4	44.0	1,936.00	200.90	940.826	4,487.6678
9	7.0	3.8	6.8	3.7	3.6	2.0	3.6	3.7	3.2	4.5	41.9	1,755.61	197.47	1,038.815	5,989.3267
10	2.1	2.8	3.1	3.8	4.2	2.7	3.6	5.1	3.4	4.0	34.8	1,211.04	127.76	492.258	1,980.2084
11	3.6	5.4	5.9	4.9	4.3	7.2	6.6	4.8	5.2	4.8	52.7	2,777.29	287.95	1,629.191	9,526.0675
12	6.7	4.0	6.7	4.1	4.8	6.7	5.2	5.0	4.7	6.1	54.0	2,916.00	301.86	1,742.214	10,343.4678
13	5.1	6.3	3.3	4.5	4.1	4.7	5.7	3.4	4.6	4.2	45.9	2,106.81	218.59	1,078.425	5,499.1639
14	3.3	2.5	2.8	1.6	4.8	3.2	1.7	3.4	2.7	3.8	29.8	888.04	97.00	339.742	1,265.0164
15	5.1	4.2	3.7	4.6	4.6	4.7	5.0	4.3	3.2	5.3	44.7	1,998.09	203.57	942.039	4,419.3509
16	7.1	6.1	6.1	6.2	5.7	5.4	7.5	5.8	6.0	5.0	60.9	3,708.81	375.81	2,350.845	14,910.5877
17	4.0	5.6	4.0	3.0	4.6	2.4	4.9	3.4	4.9	7.3	44.1	1,944.81	212.15	1,105.395	6,183.7907
18	7.3	6.6	6.0	6.2	5.8	7.6	5.7	7.7	7.5	6.9	67.3	4,529.29	458.33	3,157.039	21,980.4773
19	5.4	6.6	8.2	6.4	6.4	5.0	4.6	6.0	4.8	4.5	57.9	3,352.41	347.33	2,160.669	13,934.0897
20	7.7	7.0	6.7	7.8	7.4	5.3	5.8	7.2	5.2	6.5	66.6	4,435.56	451.64	3,112.542	21,755.8868
Total											1,008.7	53,925.39	5,581.25	33,094.825	206,937.7457

$$\langle\langle 4\rangle'\rangle' = \frac{1}{cn}\,\Sigma\Sigma x^4 = \frac{206937.7457}{200} = 1,034.6887285$$

$$\langle\langle 2\rangle'\langle 2\rangle'\rangle' = \frac{1}{cn^2}\,\Sigma(\Sigma x^2)^2 = \frac{1857865.2371}{2000} = 928.93261855$$

$$\langle\langle 1\rangle'\langle 3\rangle'\rangle' = \frac{1}{cn^2}\,\Sigma(\Sigma x)(\Sigma x^3) = \frac{1907675.5563}{2000} = 953.83777815$$

$$\langle\langle 1\rangle'\langle 1\rangle'\langle 2\rangle'\rangle' = \frac{1}{cn^3}\,\Sigma(\Sigma x)^2(\Sigma x^2) = \frac{18062136.4845}{20000} = 903.106824225$$

$$\langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle' = \frac{1}{cn^4}\,\Sigma(\Sigma x)^4 = \frac{175669499.3163}{200000} = 878.3474965815$$

<u>Sample cumulants of sample cumulants</u>

$$[[1]']' = \langle\langle 1\rangle'\rangle'$$

$$= 5.0435$$

$$[[2]']' = \frac{n}{n-1}\left(\langle\langle 2\rangle'\rangle' - \langle\langle 1\rangle'\langle 1\rangle'\rangle'\right)$$

$$= \frac{10}{9}(27.90625 - 26.962695)$$

$$= 1.04839$$

$$[[1]'[1]']' = \frac{c}{c-1}\left(\langle\langle 1\rangle'\langle 1\rangle'\rangle' - \langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\rangle'\right)$$

$$= \frac{20}{19}(26.962695 - 25.436892)$$

$$= 1.60611$$

$$[[3]']' = \frac{n^2}{(n-1)(n-2)}\left(\langle\langle 3\rangle'\rangle' - 3\langle\langle 1\rangle'\langle 2\rangle'\rangle' + 2\langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle'\right)$$

$$= \frac{100}{72}(165.474125 - 3(155.7905385) + 2(151.0792096))$$

$$= .362401$$

$$[[1]'[2]']' = \frac{cn}{(c-1)(n-1)}\left(\langle\langle 1\rangle'\langle 2\rangle'\rangle' - \langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle' - \langle\langle 1\rangle'\rangle'\langle\langle 2\rangle'\rangle'\right.$$

$$\left. + \langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 1\rangle'\rangle'\right)$$

$$= \frac{200}{171}(155.7905385 - 151.0792096 - 140.7451719 + 135.9863522)$$

$$= -\ .055545$$

$$[[1]'[1]'[1]']' = \frac{c^2}{(c-1)(c-2)}\left(\langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle' - 3\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 1\rangle'\rangle'\right.$$

$$\left. + 2\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\rangle'\right)$$

$$= \frac{400}{342}(151.0792096 - 3(135.9863522) + 2(128.2909661))$$

$$= -\ .348438$$

$$[[4]']' = \frac{n^3}{(n-1)(n-2)(n-3)}\left(\langle\langle 4\rangle'\rangle' - 3\langle\langle 2\rangle'\langle 2\rangle'\rangle' - 4\langle\langle 1\rangle'\langle 3\rangle'\rangle'\right.$$

$$\left. + 12\langle\langle 1\rangle'\langle 1\rangle'\langle 2\rangle'\rangle' - 6\langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle'\right)$$

$$+ \frac{n^2}{(n-1)(n-2)(n-3)}\left(\langle\langle 4\rangle'\rangle' + 3\langle\langle 2\rangle'\langle 2\rangle'\rangle' - 4\langle\langle 1\rangle'\langle 3\rangle'\rangle'\right)$$

$$= \frac{1000}{504}(1034.6887285 - 3(928.93261855) - 4(953.83777815)$$

$$+ 12(903.10682422) - 6(878.34749658))$$

$$+ \frac{100}{504}(1034.6887285 + 3(928.93261855)$$

$$- 4(953.83777815))$$

$$= .6948767$$

$$[[1]'[3]']' = \frac{cn^2}{(c-1)(n-1)(n-2)}\big(\langle\langle 1\rangle'\langle 3\rangle'\rangle' - \langle\langle 1\rangle'\rangle'\langle\langle 3\rangle'\rangle'$$

$$+3\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 2\rangle'\rangle' - 3\langle\langle 1\rangle'\langle 1\rangle'\langle 2\rangle'\rangle'$$

$$-2\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle' + 2\langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle'$$

$$= \frac{2000}{1368}(953.83777815 - 834.56874944 + 3(785.72958092)$$

$$- 3(903.10682422) - 2(761.96799362) + 2(878.34749658))$$

$$= -.1516010$$

$$[[1]'[1]'[2]']' = \frac{c^2n}{(c-1)(c-2)(n-1)}\big(\langle\langle 1\rangle'\langle 1\rangle'\langle 2\rangle'\rangle' - 2\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 2\rangle'\rangle'$$

$$-\langle\langle 1\rangle'\langle 1\rangle'\rangle'\langle\langle 2\rangle'\rangle' + 2\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\rangle'\langle\langle 2\rangle'\rangle'$$

$$-\langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle' + 2\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle'$$

$$+\langle\langle 1\rangle'\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 1\rangle'\rangle'$$

$$-2\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 1\rangle'\rangle'\big)$$

$$= \frac{4000}{3078}(903.10682422 - 2(785.72958092) - 752.42770734$$

$$+ 2(709.84827435) - 878.34749658$$

$$+ 2(761.96799362) + 726.98692166 - 2(685.84716748))$$

$$= -.2630524$$

$$[[1]'[1]'[1]'[1]']' = \frac{c^3}{(c-1)(c-2)(c-3)}\big(\langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle'$$

$$-3\langle\langle 1\rangle'\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 1\rangle'\rangle'$$

$$-4\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle'$$

$$+12\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 1\rangle'\rangle'$$

$$-6\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\rangle')$$

$$+\frac{c^2}{(c-1)(c-2)(c-3)}(\langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle'$$

$$+3\langle\langle 1\rangle'\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 1\rangle'\rangle'-4\langle\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle'$$

$$=\frac{8000}{5814}(878.34749658-3(726.98692166)-4(761.96799362)$$

$$+12(685.84716748)-6(647.03548734))$$

$$+\frac{400}{5814}(878.34749658+3(726.98692166)-4(761.96799362))$$

$$=-2.6974107$$

$[[2][2]']'$

$$-\frac{2}{n+1}\left\{\frac{c-1}{c}[[2]'[2]']'+[[2]']'[[2]']'-\frac{1}{n}[[4]']'\right\}$$

$$=\frac{n^2}{(n-1)}\left(\frac{1}{(c-1)(n-1)}+\frac{n-6}{(n-2)(n-3)}\right)(\langle\langle 2\rangle'\langle 2\rangle'\rangle'-2\langle\langle 1\rangle'\langle 1\rangle'\langle 2\rangle'\rangle'$$

$$+\langle\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\langle 1\rangle'\rangle')$$

$$-\frac{cn^2}{(c-1)(n-1)^2}(\langle\langle 2\rangle'\rangle'\langle\langle 2\rangle'\rangle'-2\langle\langle 1\rangle'\langle 1\rangle'\rangle'\langle\langle 2\rangle'\rangle'$$

$$+\langle\langle 1\rangle'\langle 1\rangle'\rangle'\langle\langle 1\rangle'\langle 1\rangle'\rangle')$$

$$+\frac{2n}{(n-1)(n-2)(n-3)}(\langle\langle 4\rangle'\rangle'+3\langle\langle 2\rangle'\langle 2\rangle'\rangle'-4\langle\langle 1\rangle'\langle 3\rangle'\rangle')$$

$$=\frac{100}{9}\left(\frac{1}{171}+\frac{4}{56}\right)(928.9326186-2(903.1068242)+878.3474966)$$

$$-\frac{2000}{1539}(778.7587891-2(752.4277073)+726.9869217)$$

$$+ \frac{20}{504}(1034.6887285 + 3(928.9326186) - 4(953.8377782))$$

$$= .0021893$$

Cumulant component estimates

First degree:

$[[1]']' = 5.0435 = [[1]]'$

$[1]' = [[1]]' = 5.0435$

Second degree:

$[[1]'[1]']' = 1.60611 = \frac{1}{10}[[2]]' + [[1,1]]'$

$[[2]']' = 1.04839 = [[2]]'$

$[[1,1]]' = 1.60611 - .10484 = 1.50127$

$[2]' = [[2]]' + [[1,1]]' = 2.54966$

Third degree:

$[[1]'[1]'[1]']' = -.348438 = \frac{1}{100}[[3]]' + \frac{3}{10}[[1,2]]' + [[1,1,1]]'$

$[[1]'[2]']' = -.055545 = \frac{1}{10}[[3]]' + [[1,2]]'$

$[[3]']' = .362401 = [[3]]'$

$[[1,2]]' = -.055545 - .036240 = -.091785$

$[[1,1,1]]' = -.348438 + .027536 - .003624 = -.324526$

$[3]' = [[3]]' + 3[[1,2]]' + [[1,1,1]]' = -.053910$

Fourth degree:

$$[[1]'[1]'[1]'[1]']' = -2.6974107 = \frac{1}{1000}[[4]]' + \frac{3}{100}[[2,2]]' + \frac{4}{100}[[1,3]]' + \frac{6}{10}[[1,1,2]]' + [[1,1,1,1]]'$$

$$[[1]'[1]'[2]]' = -.2630524 = \frac{1}{100}[[4]]' + \frac{1}{10}[[2,2]]' + \frac{2}{10}[[1,3]]' + [[1,1,2]]'$$

$$[[1]'[3]']' = -.1516010 = \frac{1}{10}[[4]]' + [[1,3]]'$$

$$[[2]'[2]']' - \frac{2}{n+1}\left\{\frac{c-1}{c}[[2]'[2]']' + [[2]']'[[2]']' - \frac{1}{n}[[4]']'\right\}$$

$$= .0021893 = \frac{1}{10}[[4]]' + [[2,2]]'$$

$$[[4]']' = .6948767 = [[4]]'$$

$$[[2,2]]' = .0021893 - .0694877 = -.0672984$$

$$[[1,3]]' = -.1516010 - .0694877 = -.2210887$$

$$[[1,1,2]]' = -.2630524 + .0442177 + .0067298 - .0069488 = -.2190537$$

$$[[1,1,1,1]]' = -2.6974107 + .1314322 + .0088435 + .0020190 - .0006949$$

$$= -2.5558109$$

$$[4]' = [[4]]' + 3[[2,2]]' + 4[[1,3]]' + 6[[1,1,2]]' + [[1,1,1,1]]'$$

$$= -4.2615064$$